REDUCING RISK IN CYBERSECURITY, FRAUD & REGULATORY COMPLIANCE

Cloudera Enterprise Data Cloud

SECURITY

$450 Billion: Annual Impact of Cyber Incidents

67%: Enterprises Exposed to Cyber threats

8 months: Average Time an Advanced Security Breach Goes Unnoticed

82%: Percentage of Breaches that Happened in Minutes

41%: Breaches Not Investigated

26%: Investigated, but Not Remediated

Sources:

Cisco cybersecurity report 2018;

Visual Capitalist - "The Changing Landscape of Business Risk" 2018;

Verizon Data Breach Investigations Report 2016

Protect the Business: Introduction

At its core, protecting the business is about mitigating risk; minimizing the exposure to loss. When looking at risk through the lens of business, there are many issues that organizations must navigate to avoid monetary and reputational losses. Most of these business risks manifest themselves in the form of cyber threats, fraud, and ever-looming compliance regulations.

As these risk areas have expanded over the years, they are getting increased attention. Yet they simply do not seem to be getting the additional resources they deserve. Otherwise, why else would we continue to lose $190B(1) to credit card fraud annually? How were half a billion(2) personal records able to be stolen in 2018? And now we potentially expose ourselves to losing 2 percent of revenue to a new compliance regulation.(3)

External risks are changing every day, as do the complexities that enterprises face behind the scenes to keep up with these changes. Applications continue to be added and bolted on as band-aid fixes, and emerging technologies open new threat surfaces.

Internal risks are expanding at the same time. Spear-phishing, activists, agents, rogue - or simply unaware - employees, combined with the increasing compression and miniaturization of compute and storage to thumbdrive scale can breach and launch an attack as well as exfiltrate critical high value information in minutes.

Yet the **data that can help combat risk continues to sit fragmented and stagnant across the enterprise** at great cost to the organization.

It doesn't have to be this way. There is a way to future-proof your strategy and increase your business agility in the areas of:

- **Cybersecurity**
- **Fraud Mitigation**
- **Data Science & Risk Modeling**
- **Regulatory Compliance**

This can be accomplished with significant benefits by making fundamental shifts in your data strategy related to risk management, as you have similarly leveraged and enhanced data and analytics capabilities for serving other key operational areas such as ERP, CRM, HCM and SCM.

This paper is offered for your consideration of how open source technology from Cloudera may help your organization **better leverage your data and analytics resources to protect your business.** This is in response to the increasing expansion, hybridization and complexity of IT ecosystems and modalities. While more visible discussions and proposals are commonly focused around growing your business and reducing costs, the focus of this paper is in helping you protect your business by leveraging the depth and breadth of resources from Cloudera.

Our view is that enterprises today can now safely consider a new approach in their corporate data strategy and **at a lower Total Cost of Ownership (TCO)** to address the changing threat landscape affecting responsibilities towards risk, security, governance and compliance. This is where the open source software

ecosystem and Cloudera come into play. Open source communities encourage anyone to participate and send in code, and Cloudera is one of the most dominant committers in the open source software community.

As a result, Cloudera Data Platform has proven to be the most advanced, performant, scalable and reliable open source enterprise data cloud solution available in the market today. The advantage for you is the **distribution of the overall R&D investment is made over a broader, more knowledgeable ecosystem** while at the same time **reducing exposure to vendor lock-in and technology obsolescence.** This benefits the general community, and our customers in particular.

On the surface, "open source" may imply that hackers might be able to sneak backdoors into popular open source projects. In reality, open source provides better weapons to the defense of your enterprise, through transparency and community. **The inherent strong governance process of open source projects, with many eyes reviewing the code, removes this risk.** This is where open source companies and distributions come in, guaranteeing releases with the kind of screening, penetration testing, and security analysis you would expect from a commercial software vendor but **at an amortized, distributed cost. And removing the risk of vendor lock-in at the same time**, allowing for increased agility and flexibility as your needs evolve.

Sharing code, ideas, and intelligence data is very common on the black hat side through dark web communities and market places. But white hat open source communities follow this process as well. They apply the same kind of community leverage by sharing code and intel but do so in a community of trust with a watchful eye.

With the **rapid adoption of machine learning and artificial intelligence on both sides of the battle**, the open source debate takes on a new angle. Powerful platforms from open source software are available to both sides. But the quality and efficacy of models is usually dictated by data availability for training. While the attackers lack the scruples to obey compliance and data privacy rules, and can therefore exploit more of the information, they are reliant on information leaked or stolen, while **the defense can usually benefit from a more complete picture** - if they can use the full range of their data effectively to build defensive models. As a result, you can take advantage of new ways – and **at faster, lower cost cycles - to reduce business risk** as it relates to cybersecurity, fraud, and compliance with increased agility at the same time.

Protect the Business: Cybersecurity

Cybersecurity has become an urgent topic of conversation for organizations across every industry, and a priority investment among most IT departments. And for good reason: the average breach costs $200 per lost customer record(4) and even more for lost intellectual property. **Reputational damage in and of itself can even kill a business outright.** As a result, it should not be a surprise that organizations are looking for new ways to detect and investigate cyber threats.

Attackers have become more sophisticated and the attack surfaces that can be exploited by them have expanded. **As the number of attacks have increased, organizations find themselves exposed to an onslaught of novel and previously unseen attacks.** Combined with the threat of inside rogue users and

limited availability of skilled resources for detecting and responding to these threats, it is clear organizations face an enormous challenge. The disparate and expanding choice of tools available to the **Security Operations Center (SOC) are not built for the hyperconnected world** they now operate within.

The threat landscape is changing rapidly. The number of touch points is exploding and so is the number of entry points for malicious activity. Hackers are getting more sophisticated. Activists are getting more aggressive. Agencies are getting more assertive. With traditional cybersecurity systems such as a Security Information Events Management (SIEM), organizations face **data and analytic constraints that are causing threats to go unnoticed** and data breaches to happen. SIEM cannot monitor every corner of the enterprise because of technology, human resource and economic constraints; they are hard-pressed to discover known threats until it is too late, and they only hold a subset of data that makes it difficult to use historic data for investigation and remediation. With 71 percent(5) of organizations saying it is impossible to leverage advanced analytics on traditional systems to discover advanced threats, this is **forcing organizations to rethink their cybersecurity strategy.**

Our view is that enterprises today can **safely consider a new approach** in their corporate data strategy to address the changing threat landscape affecting their cybersecurity responsibilities. That is where the open source software ecosystem and Cloudera come into play. Open source provides better weapons to the defense of your enterprise, through transparency and community. **The inherent strong governance process of open source projects, with many eyes reviewing the code, removes this risk.** This is where open source companies and distributions come in, guaranteeing releases with the kind of screening, penetration testing, and security analysis you could expect from a commercial software vendor. And removing the risk of vendor lock-in while ensuring agility and flexibility as needs evolve.

Sharing code, ideas, and intelligence data is very common on the black hat side through dark web communities and market places. White hat open source communities learn from this in the process, and apply the same kind of community leverage by sharing code and intel, but do so in a community of trust with a watchful eye.

With the move towards machine learning and artificial intelligence on both sides of the battle, the open source debate takes on a new angle. Powerful platforms from open source software are available to both sides, but the quality and efficacy of models is usually dictated by data availability for training. While the attackers lack the scruples to obey compliance and data privacy rules, and can therefore exploit more of the information, they are reliant on information leaked or stolen, while **the defense can usually benefit from a more complete picture** if they can use the full range of their data effectively to build defensive models.

The Cloudera Data Platform is built on the latest open source projects and available in a variety of form factors. The result is **a highly performant, scalable and reliable platform designed to process large** variety, volume and velocity of data that can solve for the most **complex and demanding enterprise data management requirements. This brings rise to a new class of cybersecurity solution – one designed as a modular framework** that can both incorporate existing investments and assimilate new technologies, sources, data and ML

models over time – and in real-time - **to detect previously unseen threats early in the kill chain**—helping organizations avoid financial and reputational damage.

The Cloudera Data Platform **modernizes an organization's cybersecurity architecture with the ability to combine BOTH real-time streaming machine log data AND massively scalable, low cost storage/compute at its core, AND an integrated data science/ML algorithm workbench - providing a highly scalable advanced security analytics framework.** Built with open source community software projects, CDP provides a canonical data model that offers organizations the ability **to detect cyber anomalies in real-time, focused heavily on streaming data and fast data processing at scale** to enable organizations to rapidly respond to identified anomalies **in telemetry data from most known security endpoint services**, machine generated logs, intrusion detection systems and network & threat intel feed source agents, as well as more traditional enterprise transaction systems, such as ERP, CRM, HCM applications.

The result is the ability to **detect advanced threats 2.25 times faster(6)** and accelerate threat mitigation leveraging big data and advanced analytics (machine learning, predictive analytics, etc.).

Unlike traditional solutions that provide signature and correlation analysis across subsets of security data, the open source-based Cloudera Cybersecurity Platform can ingest, store, process, and analyze any volume of data with any analytic type. Having access to all the raw data in one place can **help uncover new insights and patterns. This allows for behavior-driven analytics that can detect the smallest changes in user or system behavior**—traditionally the most reliable indicators of compromise.

Integrating existing cyber defenses, Cloudera Cybersecurity Platform allows organizations to quickly deploy and improve their security posture with no disruption.

Protect the Business: Fraud Mitigation

Challenges in fraud detection have increased dramatically with the introduction of new access points to service offerings and increased sophistication of perpetrators. Furthermore, as companies expand into new markets, they face new fraud risks that must be modeled.

With each new access point, firms are more **susceptible to new methodologies and ever more complex cross-channel fraud.** Let's take money laundering as an example. Unlike other forms of fraud that are identified with machine learning algorithms that detect anomalies and outliers, money laundering schemes are designed to closely mimic typical banking behaviors—hiding clandestine or illegal activities behind a taxable, auditable mainstream business front—and are, therefore, characteristically less anomalous.

The thresholds mandated by reporting policies like Bank Secrecy Act (BSA) and Dodd-Frank utilized by first- and second-generation anti-money laundering (AML) systems are well known, so criminals have little difficulty modeling the source of their above-board trade and transaction behaviors to be largely imperceptible, even to specialized software. **As a result, these systems must be enriched with much larger and more diverse data sets to isolate signals of possible money laundering.**

The Cloudera Data Platform is built on the latest open source projects and

available in a variety of form factors. The result is a highly performant, scalable and reliable platform designed to process large variety, volume and velocity of data that can solve for the most complex and demanding enterprise data management requirements. This brings rise to **a new class of fraud detection & prevention solution – one designed as a modular framework** that can both incorporate existing investments and assimilate new technologies, sources, data and ML models over time – and in real-time - **to detect previously unseen transactions early in the fraud scheme**—helping organizations avoid financial loss and regulatory damage.

A Cloudera Data Platform is the ideal platform for AML because it both incorporates and augments all existing core functions of new or specialized systems to better handle the volume, variety and velocity of big data: data collection, data preparation, automated evaluation, model building, and investigation. A modern AML architecture based on a Cloudera Data Platform starts with a fully integrated open source-based enterprise data hub, with ingestion and staging capabilities that **can support streaming massive complex data in real-time, and a centralized, standardized Data Science Workbench** as needed. This allows for legacy solutions to provide faster run times for the predictive models **and perform the actual fraud detection in real-time more efficiently.**

Beyond the introductory use case of more expansive and affordable storage, Cloudera Data Platform's **natural fit for back testing against long-term descriptive data is gaining popularity for more advanced AML workloads.** Additionally, the availability of other Data Science and Machine Learning components in the Cloudera stack for exploration, discovery, investigation, and forensics can be leveraged in a fraction of the time it takes to stand up a more traditional Data Science project.

Protect the Business: Data Science & Risk Modeling

More data-driven risk modeling is needed to address the rapidly evolving types of fraud, crime and related government-inspired regulations in this changing landscape. For example, it was not too long ago that "Privacy" was NOT considered a risk issue for an enterprise beyond keeping employee data private. Today: We have GDPR and nation-state sponsored cyberwars that include fraud, theft and … more. So, too, not that long-ago people would never have considered depositing a check or transferring money on/from their cell phone. The introduction of the iPhone was not that long ago, and in that time: electronic banking from your couch via your phone is considered common by most people.

Risk management is heavily dependent on modeling & analysis processes. As a result, it is a persistent compute hungry and data intensive activity that crosses multiple functional silos. This creates many challenges for any organization, but particularly for firms in financial services, insurance, telecommunications, energy and life sciences that are extremely data driven enterprises by nature.

As such, it is no longer enough to rely on brilliant quant staff with complex algorithms driven by data samples, which leaves firms susceptible to hidden deficiencies and irreconcilable predictions. **The most valuable tools help firms identify and mitigate areas of risk such as fraud, theft and regulatory exposure fundamentally require the ingestion, processing and analysis of detailed traces across all operational systems.** The volume, velocity and

variety of that data is very big indeed. Truly Big Data.

Since perpetrators work hard to exploit gaps in financial or billing systems, firms must be vigilant and self-aware of every place where they may be exposed.

Our view is that enterprises today can safely consider a new approach in their corporate data strategy to address the changing threat landscape affecting their risk management responsibilities that new technologies and applications are introducing to their business applications. That is where the open source software ecosystem and the Cloudera Data Platform come into play.

The Cloudera Data Platform is built on the latest open source projects and available in a variety of form factors. The result is a **highly performant, scalable and reliable platform designed to process large variety, volume and velocity of data** that can solve for the most complex and demanding enterprise data management requirements.

This introduces a higher level of risk modeling & analysis capabilities with Cloudera Data Science Workbench (CDSW). **CDSW is designed as a modular framework that can both incorporate existing investments and assimilate new technologies, sources, data and ML models** over time – and in real-time.

With Python, R, and Scala directly in the web browser, CDSW delivers a self-service experience data scientists will love. Download and experiment with the latest libraries and frameworks in customizable project environments that work just like your laptop. Access any data, anywhere—from cloud object storage to data warehouses, **CDSW provides connectivity not only to all Cloudera Data Platform options, but also to the systems your data science teams rely on for analysis.** This enables your teams to detect previously unseen risks to potential fraud, theft or regulatory exposure earlier in the chain—helping organizations avoid financial and reputational damage.

A modern Risk Modeling & Analysis architecture based on a Cloudera Data Platform in concert with Cloudera Data Science Workbench starts with a fully integrated open source-based enterprise data hub, with ingestion and staging capabilities that can support streaming massive complex data in real-time, as needed. This allows for legacy solutions to provide faster run times for the predictive models and perform the actual risk modeling and detection in real-time, as needed.

The ability to collect and analyze detailed behaviors from online channels and automated systems as well as existing data stores at scale with Cloudera offers a significant advantage. **Using Cloudera solutions, risk modeling & analysis for fraud, theft detection and regulatory compliance teams can combine logically linked accounts by looking for common patterns of money movement and related transactions.** Like the way social networking companies find relationship links that are complicated to identify, antifraud teams search for connections that are implied by detailed trace data.

Collecting detailed information on both customer and internal interactions leads to **new models that help identify patterns of normal and suspect behavior.** Advanced risk detection needs advanced analytics, including data mining and modeling techniques that leverage advances in ML. Cloudera brings the processing power to analyze massive amounts of data to quickly identity and prevent fraud.

Protect the Business: Regulatory Compliance

The cost and complexity of compliance for organizations have escalated significantly in recent years. Stringent regulatory compliance laws have been put in place to improve operational transparency, digital privacy, and customer protection. **Organizations are held much more accountable** for their actions and are required to be able to access years of historical data in response to regulators' requests for information at any given time.

For example, the Dodd-Frank Act requires firms to maintain records for at least five years; Basel guidelines mandate retention of risk and transaction data for three to five years; and Sarbanes-Oxley requires firms to maintain audit work papers and required information for a minimum of seven years. These **records must be available on demand, or in some cases must be normalized and sent to regulators proactively**. Increasingly, organizations are setting up internal audits to identify and prevent rogue employee behavior or loopholes in internal systems or processes.

And there's **GDPR. HIPAA. CASL. Coming soon: EU Cybersecurity Act.** Tracking and complying with the ever-increasing regulatory state is only one part of the cost; an even greater cost is the possible risk of fines and public relations impacts if something is missed or lost.

Frequent changes in regulations have tested the ability of legacy compliance systems to respond in a timely manner. Partly because of these pressures, **leading companies have realized that the key to optimizing their business operations is in maintaining an efficient and large-scale data management infrastructure.** This is very expensive and complex to accommodate using traditional systems.

Cloudera helps organizations create a secure, auditable, and searchable unified repository for all the data at a fraction of the cost of traditional systems. This platform can be used for multiple risk and compliance use cases: PCI, HIPAA, VAR, Monte Carlo simulations, CCAR, Basel III, Solvency II, BCBS239, MiFID II, FRTB, etc. New data sources can be added easily and changes can be implemented more quickly. By eliminating data silos and time-consuming processes such as ETL/ELT, the time it takes to prepare the data for analysis can be drastically reduced.

How would I get started?

We are here to help. The first step is to arrange for an initial meeting with our team to understand what you're trying to accomplish, and identify business-impacting use cases. After that, most of our customers prefer to do some form of proof-of-concept to get a sense of the power of our solutions.

In some cases, we'll recommend one of our systems integration partners, or perhaps you'd rather engage directly with Cloudera's professional services team —either way is fine with us. And, whether you want to deploy on-premise or in the cloud, we will support you 100 percent.

Lastly, to help ensure the likelihood of success, we strongly recommend you provide your team with training. In fact, Cloudera provides comprehensive and in-depth training for all your opensource software needs, and for various roles within the organization. In addition to making sure you get the most out of your

Cloudera deployment, our training will help you with recruiting and retention of the DevOps people essential to open source software.

Summary

The risk landscape is changing rapidly. Future-proofing your risk strategy by using open source technologies to leverage data and analytic more effectively is key. Unifying data, properly securing appropriate access, and applying advanced analytics and machine learning technologies against it will allow you to stay one step ahead of the looming risk threats, instead of reacting to continuous tactical challenges. Cloudera is here to help you reduce your overall risk exposure by more effectively leveraging your data.

About Cloudera

Cloudera delivers the modern platform for data management and analytics for the enterprise data cloud. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Data Platform —the fastest, easiest, and most secure data platform built on open source technology. Our customers can efficiently capture, store, process, and analyze vast amounts of data—empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training, and professional services.

References

1 "Solving the $190 Billion Annual Fraud Problem: More on Jumio"—Forbes. 12 Oct. 2016
http://www.forbes.com/sites/haydnshaughnessy/2011/03/24/solving-the-190-billion-annual-fraud-scam-more-on-jumio/1

2 "You've been breached: Hackers stole nearly half a billion personal records in 2018"—NBC News 4 Feb. 2019
https://www.nbcnews.com/business/consumer/you-ve-been-breached-hackers-stole-nearly-half-billion-personal-n966496
3 "General Data Protection Regulation"—Wikipedia, the free encyclopedia. 12 June 2019 https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

4 Data Breach Costs Top $200 Per Customer Record"—CIO. 2016. 12 Oct. 2016
https://www.cio.com/article/2421114/security0/data-breach-costs-top--200-per-customer-record.html

5 "Big Data Cybersecurity Analytics Research Report"—Ponemon Research Institute. August.2016
- https://www.cloudera.com/content/dam/www/marketing/resources/analyst-reports/big-data-cybersecurity-analytics-research-report.pdf.landing.html

6 "Big Data Cybersecurity Analytics Research Report"—Ponemon Research Institute. August.2016 -
https://www.cloudera.com/content/dam/www/marketing/resources/analyst-reports/big-data-cybersecurity-analytics-research-report.pdf.landing.html

Cloudera, Inc. 395 Page Mill Road Palo Alto, CA 94306 USA cloudera.com